Bayesian multivariate probability of success with strict control of type I error

Ethan M. Alt & Matthew A. Psioda*

April 15, 2021

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

*joint work with Joseph G. Ibrahim

Table of contents I

Case study: the COMPASS study (Duncan *et al.* 2017)

- Two-arm, cluster randomized pragmatic trial (clusters ignored in data application)
- Treatment: novel post-acute stroke care model (eCareplan)
- Endpoints of interest:
  1. Stroke impact scale (SIS-16)
  2. Self-rated health
  3. PROMIS global health scale (Hays *et al.* 2009)
- Key question: what sample size yields a design with a high probability of clinical success for a future trial to hit on multiple endpoints?

Introduction

- Increased interest among practitioners in computing the probability of having a successful clinical trial
- Framework: Chuang-Stein (*Pharmaceutical Statistics*, 2006)
- Standard methods to compute sample size rely on *statistical power*
  - Power is a *conditional value*
  - Power is *not* the probability of a successful clinical trial
- Probability of success (POS) may be defined as the expected value of power with respect to a specified distribution for the effect size:

$$\text{POS} = \int P(\text{Trial meets success critera}|\Delta)p(\Delta|D)d\Delta$$

POS for linear model

- Ibrahim and others (2015) extended POS to the univariate linear model in the presence of historical data $D_{0k} = \{(y_{0ki}, z_{0ki}, \boldsymbol{x}_{0ki}), i = 1, \ldots, n_k\}$, $k = 1, 2$.

$$\text{POS} = \int P(\text{success}|z, \boldsymbol{x}, \boldsymbol{\theta}) f(z) f(\boldsymbol{x}|\boldsymbol{\alpha}) \pi^{(v)}(\boldsymbol{\theta}, \boldsymbol{\alpha}) dz d\boldsymbol{x} d\boldsymbol{\theta} d\boldsymbol{\alpha},$$

where

$$P(\text{success}|z, \boldsymbol{x}, \boldsymbol{\theta}) = E\left[1\{P(\beta_1 > TV|D, \pi^{(f)}) \geq \gamma\}|z, \boldsymbol{x}, \boldsymbol{\theta}\right]$$

- The quantity TV is the "target value"
- The prior $\pi^{(v)}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ is referred to as the *validation*, *sampling*, or *design* prior
- The prior $\pi^{(f)}(\boldsymbol{\theta})$ is referred to as the *fitting* or *analysis* prior

The validation and fitting priors

- The validation prior is specified as $\pi^{(v)}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \pi_1^{(v)}(\boldsymbol{\theta})\pi_2^{(v)}(\boldsymbol{\alpha})$ where we utilize the power prior (Ibrahim, 2000)

$$\pi_1^{(v)}(\boldsymbol{\theta}) = [L(\boldsymbol{\theta}|D_{01})]^{a_{01}}\, \pi_{10}^{(v)}(\boldsymbol{\theta}),$$

$$\pi_2^{(v)}(\boldsymbol{\alpha}) = \left[\prod_{k=1}^{2} L(\boldsymbol{\alpha}|D_{0k})^{b_{0k}}\right] \pi_{20}^{(v)}(\boldsymbol{\alpha})$$

- The likelihood for the covariate parameters is obtained via factorization

$$L(\boldsymbol{\alpha}|D_{0k}) = \prod_{j=1}^{p}\prod_{i=1}^{n_{0k}} f(x_{ij}|x_{i1}, \ldots, x_{i,j-1}; \boldsymbol{\alpha}_j)$$

- The fitting prior is specified as

$$\pi^{(f)}(\boldsymbol{\theta}) = [L(\boldsymbol{\theta}|D_{02})]^{a_{02}}\, \pi_0^{(f)}(\boldsymbol{\theta})$$

Computation of POS

- Given samples $\{(\boldsymbol{\theta}^{(m)}, \boldsymbol{\alpha}^{(m)}), m = 1, \ldots, M\}$ and a sample size n, future data sets $D^{(m)} = \{(y_i^{(m)}, z_i^{(m)}, \boldsymbol{x}_i^{(m)}), i = 1, \ldots, n\}$ may be simulated via the prior predictive distribution for the treatment variable, covariates, and outcomes

- For each future data set $D^{(m)}$, the posterior density of the treatment effect, $p(\beta_1|D^{(m)})$, may be obtained

- POS may be estimated as

$$\text{POS} = \frac{1}{M}\sum_{m=1}^{M} 1\{P(\beta_1 > TV|D^{(m)}) \geq \gamma\},$$

where $P(\beta_1 > TV|D^{(m)})$ is the posterior probability that the success criterion is satisfied based on $D^{(m)}$

Seemingly Unrelated Regression (SUR)

SUR Model

$$
\begin{aligned}
\boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{u}_i \\
\boldsymbol{y}_i &= (y_{i1}, \ldots, y_{iJ})' \in \mathbb{R}^J \\
\boldsymbol{X}_i &= \text{blkdiag}\left\{\boldsymbol{x}_{i1}', \ldots, \boldsymbol{x}_{iJ}'\right\} \in \mathbb{R}^{J \times p}, \qquad p = \textstyle\sum_{j=1}^{J} p_j \\
\boldsymbol{\beta} &= (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_J')' \in \mathbb{R}^p \\
\boldsymbol{u}_i &\sim N_J(\mathbf{0}, \boldsymbol{\Sigma}), \qquad \boldsymbol{\Sigma} \in \mathbb{R}^{J \times J}
\end{aligned}
$$

- Most general multivariate normal linear model
- Allows each response to have its own set of covariates

Bayesian analysis of the SUR model

- The likelihood for a SUR model may be written as

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{y}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})' \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n\right)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\}$$

  where $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J)'$, $\boldsymbol{X} = \text{blkdiag}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J\}$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)'$

- Bayesian analysis requires prior specification for $\boldsymbol{\beta}, \boldsymbol{\Sigma}$

- The following prior is noninformative and enables us to obtain samples via direct Monte Carlo (DMC) (Zellner and Ando 2010)

$$\pi(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(J+1)/2}$$

What is “success”

- Ibrahim *et al.* 2015, studying the univariate POS problem, defined success as

$$\text{success} = 1\{\beta > TV\}$$

- For multivariate POS, we utilize the more general definition of success

$$\text{success} = 1\{\boldsymbol{\beta} \in \Omega\},$$

  where Ω is a set that defines how success is achieved.

- We may be interested in several different specifications of Ω

$\Omega = \{\boldsymbol{\beta} : \beta_1 > TV_1\}$	success in a sole primary endpoint
$\Omega = \{\boldsymbol{\beta} : \cap_{j=1}^{J}\{\beta_j > TV_j\}\}$	co-primary endpoints
$\Omega = \{\boldsymbol{\beta} : \cup_{j=1}^{J}\{\beta_j > TV_j\}\}$	multiple primary endpoints

The proposed method

- Let $\theta = (\beta, \Sigma)$. Multivariate POS is expressed mathematically as

$$\text{POS} = \int P(\text{success}|z, \boldsymbol{x}, \theta) f(z) f(\boldsymbol{x}|\alpha) \pi^{(v)}(\theta, \alpha) dz d\boldsymbol{x} d\theta d\alpha,$$

where $f(z)$ is the known distribution for the treatment effect, $f(\boldsymbol{x}|\alpha)$ is the density of the covariates, $\pi^{(v)}$ is a *validation prior*, and

$$P(\text{success}|z, x, \theta) = E\left[1\left\{P(\beta \in \Omega|D, \pi^{(f)}) \geq \gamma\right\}|z, \boldsymbol{x}, \theta\right]$$

- $P(\beta \in \Omega|D, \pi^{(f)})$ is the posterior probability that β lies within the region of success (Ω) given the future data (D) with respect to the fitting prior $\pi^{(f)}(\theta)$

Prior elicitation

- Suppose we possess recent historical data D_{01} and possibly an older historical data set D_{02}
- We specify the validation prior $\pi^{(v)}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \pi_1^{(v)}(\boldsymbol{\theta})\pi_2^{(v)}(\boldsymbol{\alpha})$, where

$$\pi_1^{(v)}(\boldsymbol{\theta}) \propto L(\boldsymbol{\theta}|D_{01})|\boldsymbol{\Sigma}|^{-(J+1)/2}$$

$$\pi_2^{(v)}(\boldsymbol{\alpha}) \propto \left(\prod_{k=1}^{2}[f(\boldsymbol{x}|\boldsymbol{\alpha})]^{a_{0k}}\right)\pi_{20}^{(v)}(\boldsymbol{\alpha})$$

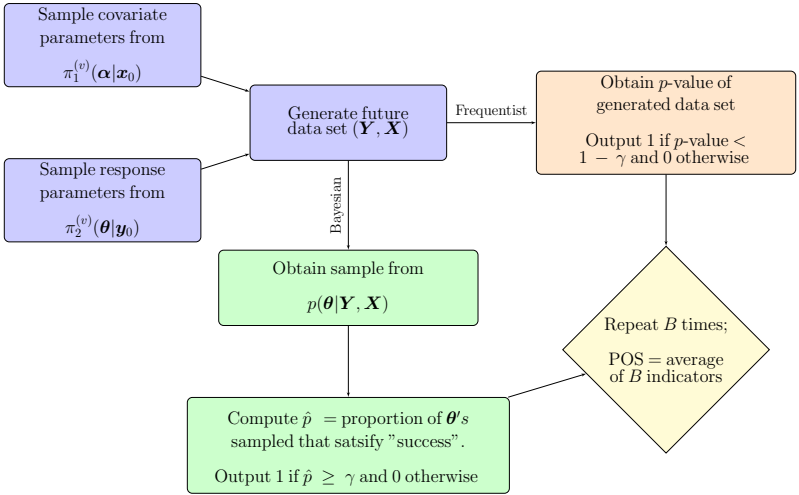
- The fitting prior is specified as a power prior

$$\pi^{(f)}(\boldsymbol{\theta}) \propto [L(\boldsymbol{\theta}|D_{02})]^{b_{02}}\,|\boldsymbol{\Sigma}|^{-(J+1)/2}$$

The proposed method (3)

- Ibrahim *et al.* 2015 propose to group the variables according to distribution (e.g., Gaussian, binomial, etc.), and generate in that order
- We propose to specify a hierarchy of generation order
- Suppose x_1 = gender, x_2 = weight, and x_3 = tumor size. Seems reasonable to generate $f(\boldsymbol{x})$ by $f(x_1)f(x_2|x_1)f(x_3|x_2, x_1)$
- We assume covariates are generated from a GLM. Suppose the first K_1 of K covariates have a dispersion parameter. We specify the initial prior $\pi_{20}^{(v)}(\alpha) = \prod_{k=1}^{K_1} \text{Gamma}(\varphi_k|\alpha_0, \gamma_0)$ where, for our data analysis and simulations, we take $\alpha_0 = \gamma_0 = 0.1$

POS Algorithm



Type I error control

- Suppose $\Omega = \cup_{j=1}^{2}\{\beta_{1j} > 0\}$ and $\overline{\Omega} = \cap_{j=1}^{2}\{\beta_{1j} = 0\}$. We augment the validation prior to be

$$\tilde{\pi}_1^{(v)}(\boldsymbol{\theta}) = \pi_1^{(v)}(\boldsymbol{\theta}) I(\boldsymbol{\theta} \in \overline{\Omega}),$$

  so that our future data sets are being generated based on a point mass at 0 for each of the treatment effects of interest

- $\mathrm{POS}(\{\beta_{1j} > 0\}) = 1 - \gamma := \alpha$ for $j = 1, 2$, and type I error control is established

- However,

$$\begin{aligned}\mathrm{POS}(\Omega) &= \mathrm{POS}(\{\beta_{11} > 0\}) + \mathrm{POS}(\{\beta_{12} > 0\}) - \mathrm{POS}(\cap_{j=1}^{2}\{\beta_{1j} > 0\}) \\ &= \alpha + \alpha - \alpha^*\end{aligned}$$

- FWER control is established if and only if $\alpha^* \geq \alpha$, i.e., if and only if $\mathrm{Corr}(\beta_{1j}, \beta_{12}) = 1$.

Type I error control (2)

- Consider replacing $\text{POS}(\cap_{j=1}^{2}\{\beta_{1j} > 0\})$ with

$$\text{POS}^*(\cap_{j=1}^{2}\{\beta_{1j} > 0\}) = \max\{\text{POS}(\cap_{j=1}^{2}\{\beta_{1j} > 0\}), \alpha\}$$

then FWER is controlled at exactly level α. This leads to the following theorem:

Theorem

Let $\Omega = \cup_{j=1}^{K}\{\beta_{1j} > 0\}$ *for some* $1 \leq K \leq J$. *Then*

1. $\text{POS}(\Omega) \geq \alpha$ *with strict inequality holding whenever* $\rho_{jk} := \text{Corr}(\beta_{1j}, \beta_{1k}) \neq 1$ *for any* $j \neq k$.
2. $\text{POS}^*(\Omega) = \alpha$ *for any* ρ_{jk}.

We can also write

$$\text{POS}^*(\Omega) = \text{POS}(\Omega) - \max\{\alpha - \text{POS}(\cap_{j=1}^{2}\{\beta_{1j} > 0\}), 0\}$$

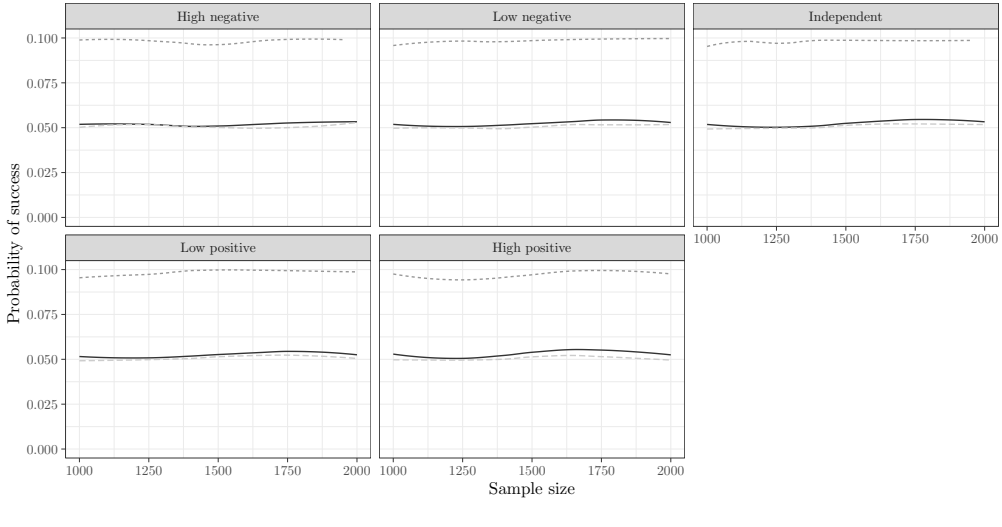
Simulation study

- Simulated historical data are based on summary statistics of the COMPASS study (except variances were halved)
- Correlations considered were

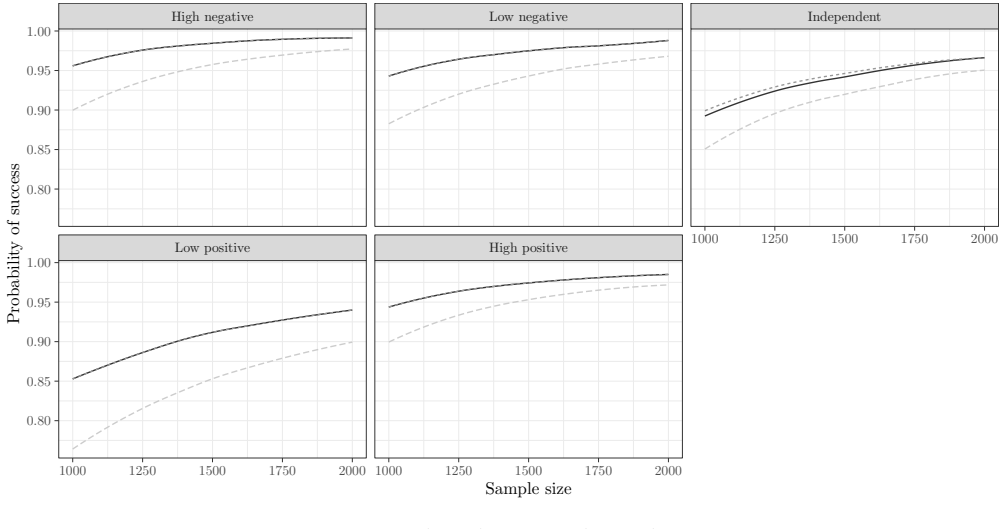
$$(\rho_{12}, \rho_{13}, \rho_{23}) \in \{(-0.3, -0.4, -0.7), (-0.05, -0.1, -0.2), (0, 0, 0), (0.05, 0.1, 0.2), (0.3, 0.4, 0.7)\}$$

- Negative of PROMIS score taken to make all treatment effects positive to make more sense of the correlation effect
- $\beta_1 = (0.0333, 0.1667, 0.5980)$ for SIS-16, self-rated health, and PROMIS score, respectively

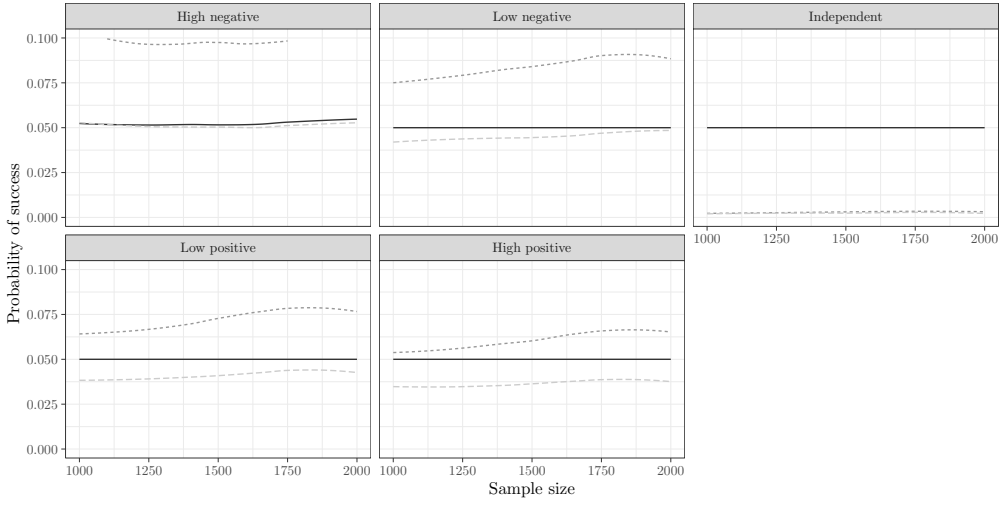
$\{\beta_{11} > 0 \cup \beta_{12} > 0\}$: Type I Error



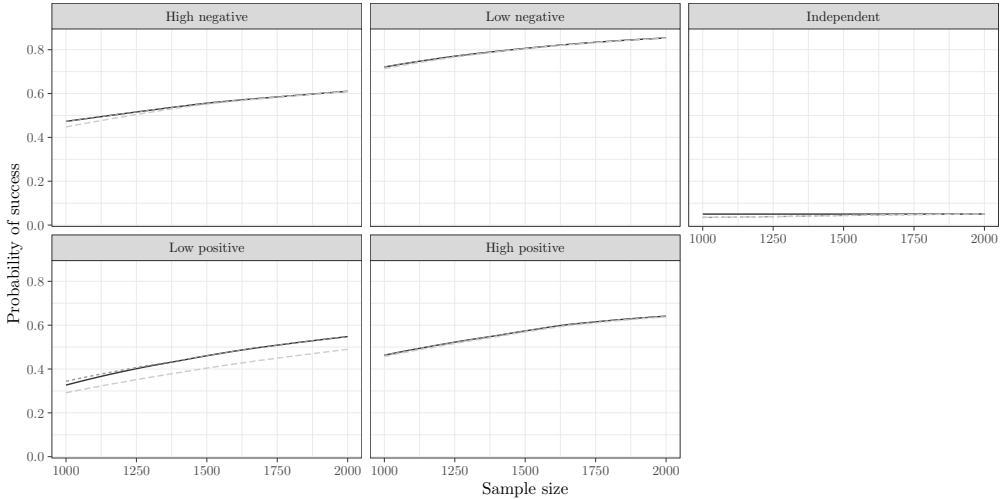
$\{\beta_{11} > 0 \cup \beta_{12} > 0\}$: BCEP



$\{\beta_{11} > 0 \cap (\beta_{12} > 0 \cup \beta_{13} > 0)\}$: Type I error



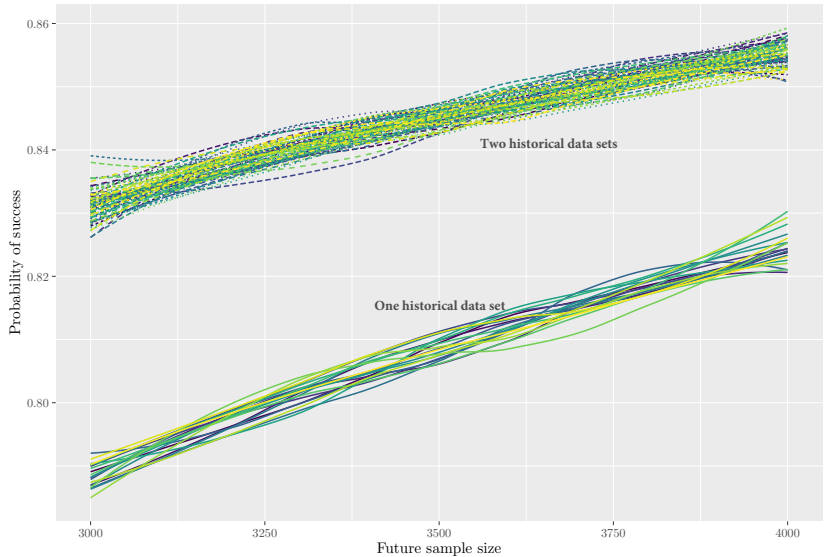
$\{\beta_{11} > 0 \cap (\beta_{12} > 0 \cup \beta_{13} > 0)\}$: BCEP



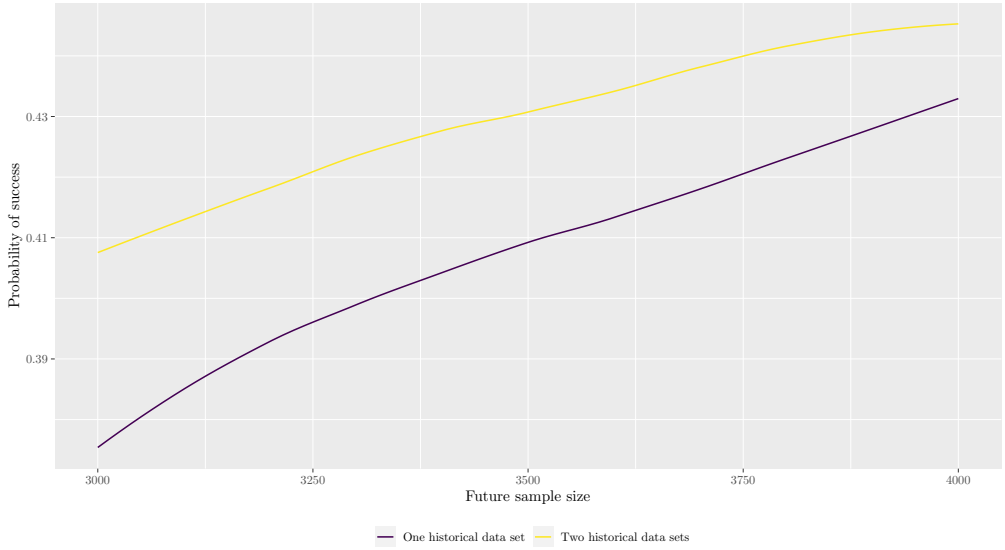
Data application - COMPASS study

- Pilot data were considered older historical data D_{02}
- Non-pilot data were utilized as newer historical data D_{01}
- Controls utilized were stroke history, TIA history, linear and quadratic terms of age, race (white or non-white), severity of stroke, whether the patient had insurance (to control for socioeconomic status), and a binary variable indicating whether the patient was hospitalized due to a stroke or a transient ischemic attack (TIA)
- Log-transform utilized for SIS-16 variable due to asymmetry and skew
- The following plots report POS for various definitions of Ω

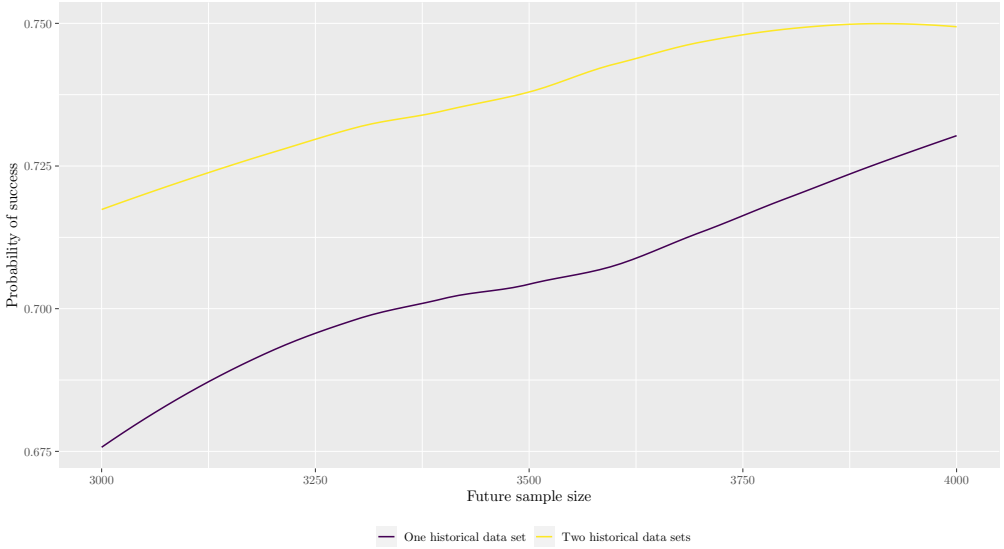
Primary Endpoint: SIS-16



Complete success: $\Omega = \cap_{j=1}^{3}\{\beta_j > 0\}$



SIS-16 and at least one of the two secondary



A note on small historical data sets

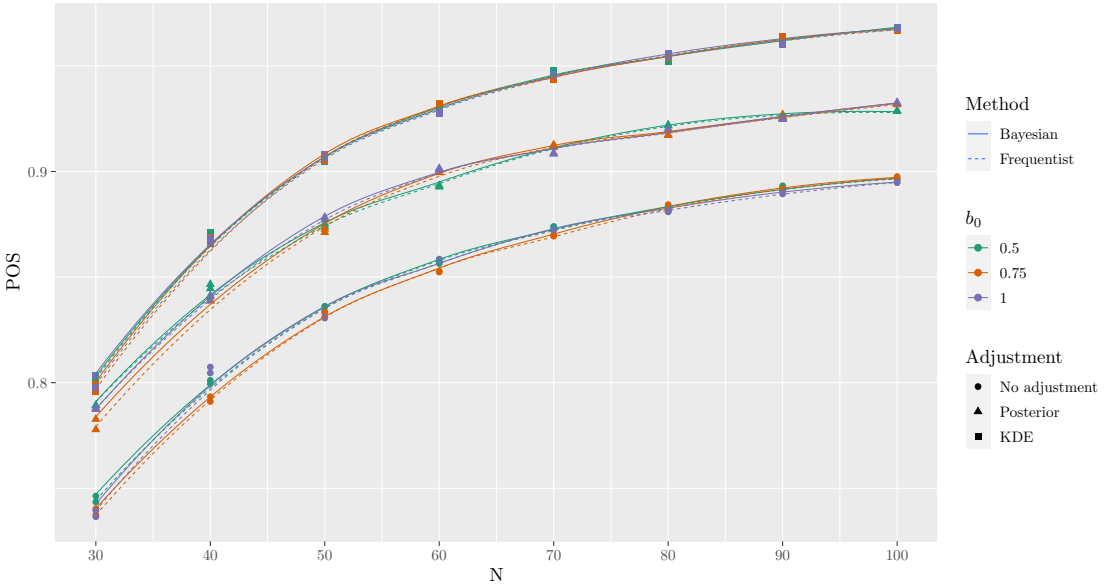
- For rare diseases, Phase II sample sizes are typically very small
- Implausible treatment effects may be sampled if the sample size is too low
- There are several possible adjustments one can make:
  1. Restrict samples for efficacy in the treatment effect: *Bayesian conditional expected power*
  2. Use informative priors for treatment effects
  3. Restrict samples of treatment effects to the q^{th} highest posterior density (HPD) region for some $0 < q < 1$
- We focus on (3) and propose two different mechanisms:
  1. HPD region of all parameters
  2. HPD region of only treatment effects estimated using KDE

Rare disease data setting

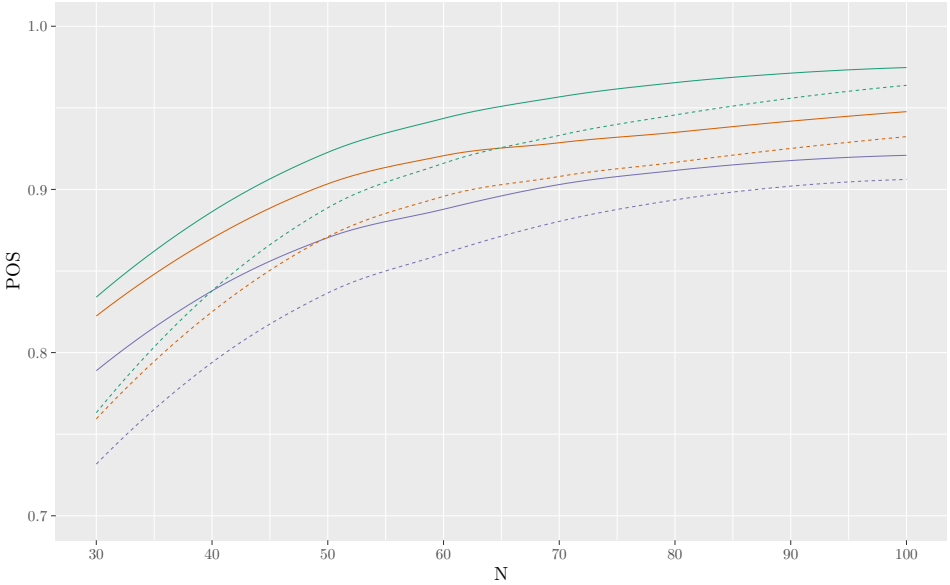
- Phase 2 trial of Ivacaftor in subjects with cystic fibrosis (Vertex Pharmaceuticals, 2007-12)
- Phase 2 data suggested treatment efficacious, but sample size was very small
- Phase 3 trial conducted 2012-15:
  - **Primary endpoint**: Absolute change from baseline in percent predicted forced expiratory volume in 1 Second (FEV1)
  - **Key secondary endpoints:**
    - Change from baseline in sweat chloride
    - Change from baseline in Cystic Fibrosis Questionnaire-Revised (CFQ-R) respiratory domain score

The simulated Phase II data

- Phase II data sample size was $n_0 = 16$ (8 treatment, 8 control)
- $E(\Delta y_{ij}) = \beta_{1j} z_i + \boldsymbol{x}'_{ij} \beta_{2j}$
- $\beta_1 = (6.4, 3.5, -49.1)'$
- $\text{diag}(\Sigma) = (5.12, 7.05, 12.27)'$
- $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.25, -0.25, -0.33)$
- For all outcomes, $\boldsymbol{x}_{ij}$ includes an intercept term, linear and square terms of age, weight, BMI, and sex
- Baseline levels for FEV1 and CFQ-R score are controlled for in their regressions
- Power computations yielded a future sample size of $n = 22$ to attain 90% power in the primary endpoint


POS for Δ FEV1
POS
0.9
0.8
30
40
50
60
70
80
90
100
N
Method
Bayesian
Frequentist
b_0
0.5
0.75
1
Adjustment
No adjustment
Posterior
KDE

POS for Δ FEV1 or Δ Sweat



Gaussian copula regression model

Gaussian copula regression model

$$(Z_{i1}, \ldots, Z_{iJ})' \sim N_J(\mathbf{0}, \boldsymbol{\Gamma})$$
$$Y_{ij} = F_j^{-1}(\Phi(Z_{ij})|\boldsymbol{\theta}_j, \boldsymbol{x}_{ij}), \quad \Phi = \text{standard normal CDF}$$

- If Y_{ij} is continuous, Z_{ij} is not latent given $\boldsymbol{\theta}_j, \boldsymbol{x}_{ij}$
- If Y_{ij} is discrete, Z_{ij} is latent and must be generated
- We assume $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j, \varphi_j)$, where $\varphi_j = 1$ is known in some cases
- After specifying priors for $\boldsymbol{\theta}$, $\boldsymbol{\Gamma}$ and obtaining samples, multivariate POS methods developed still applicable

Conclusion

- Only known method for multivariate POS
  - Methods of Ibrahim *et al.* 2015 and Chuang-Stein 2006 become special cases
- Asymptotically provides exact type I error control for complex hypotheses
  - Simulations suggest uniformly more powerful than Bonferroni-Holm
  - Invariant to order of testing and correlation of tests
- Unifies hypothesis testing for simple and composite hypotheses
- Extension being developed for multivariate GLM using Gaussian copula approach

References I

- C. Chuang-Stein, en, *Pharmaceutical Statistics* **5**, 305–309 (Oct. 2006).
- P. W. Duncan *et al.*, *BMC Neurology* **17** (July 2017).
- R. D. Hays, J. B. Bjorner, D. A. Revicki, K. L. Spritzer, D. Cella, *Quality of Life Research* **18**, 873–880 (Sept. 2009).
- J. G. Ibrahim, M.-H. Chen, M. Lakshminarayanan, G. F. Liu, J. F. Heyse, en, *Statistics in Medicine* **34**, 249–264 (Jan. 2015).
- A. Zellner, T. Ando, en, *Journal of Econometrics* **159**, 33–45 (Nov. 2010).